

LEEC: A Legal Element Extraction Dataset with an Extensive Domain-Specific Label System

Zongyue Xue*

xuezy21@mails.tsinghua.edu.cn
Law School, Tsinghua University
Beijing, China

Huanghai Liu*

liuhh23@mails.tsinghua.edu.cn
Law School, Tsinghua University
Beijing, China

Yiran Hu*[†]

huyr21@mails.tsinghua.edu.cn
Law School, Tsinghua University
Beijing, China

Kangle Kong

kkl22@mails.tsinghua.edu.cn
Law School, Tsinghua University
Beijing, China

Chenlu Wang

wangchenlu2014@gmail.com
Law School, Tsinghua University
Beijing, China

Yun Liu[†]

liuyun89@tsinghua.edu.cn
Law School, Tsinghua University
Beijing, China

Weixing Shen[†]

wxshen@tsinghua.edu.cn
Law School, Tsinghua University
Beijing, China

ABSTRACT

As a pivotal task in natural language processing, element extraction has gained significance in the legal domain. Extracting legal elements from judicial documents helps enhance interpretative and analytical capacities of legal cases, and thereby facilitating a wide array of downstream applications in various domains of law. Yet existing element extraction datasets are limited by their restricted access to legal knowledge and insufficient coverage of labels. To address this shortfall, we introduce a more comprehensive, large-scale criminal element extraction dataset, comprising 15,831 judicial documents and 159 labels. This dataset was constructed through two main steps: first, designing the label system by our team of legal experts based on prior legal research which identified critical factors driving and processes generating sentencing outcomes in criminal cases; second, employing the legal knowledge to annotate judicial documents according to the label system and annotation guideline. The Legal Element ExtraCtion dataset (LEEC) represents the most extensive and domain-specific legal element extraction dataset for the Chinese legal system. Leveraging the annotated data, we employed various SOTA models that validates the applicability of LEEC for Document Event Extraction (DEE) task. The LEEC dataset is available on <https://github.com/THUlawtech/LEEC>.

KEYWORDS

domain-specific, element extraction, legal information retrieval

1 INTRODUCTION

Extracting key elements of judicial documents and their relations is valuable for analyzing legal cases and making sentencing decisions. Meanwhile, the disparity between “law in books” and “law in action” introduces a significant difficulty in fully capturing the important elements in judicial practice, thereby augmenting the

complexity of element extraction. With the help of the extensive label system constructed on the legal knowledge graph by our team of legal experts, Legal Element ExtraCtion (LEEC) dataset aims to provide element mentions, trigger words and values manually annotated from large-scale judicial documents with high quality. This could facilitate automatic extraction of elements, benefiting numerous LegalAI applications, such as Legal Judgement Prediction and Similar Case Retrieval, as well as empirical legal research. Meanwhile, with an extensive label system based on prior empirical legal research, LEEC could provide comprehensive labels that are important in judicial practice yet neglected by prior studies regarding element extraction, while also contribute to the replication and innovation in empirical studies.

Inspired by the success of general-domain element extraction[6][13][5], previous studies[19][3][20] attempted to construct an element extraction system in the legal domain, leveraging both hand-crafted features and neural networks. For instance, LeCaRD[15], the first Legal Case Retrieval Dataset in China contains 107 query cases and 10,700 candidate cases selected from over 43,000 Chinese criminal judgements, was constructed. LEVEN is a large-scale Chinese Legal event detection dataset[31], with 8,116 legal documents and 150,977 human annotated event mentions in 108 event types. At present, the existing datasets in China also includes CAIL[28], Criminal[15], CJO¹, PKU², etc. However, there are several main challenges in the existing work:

(1) **Incomprehensive Label System.** Existing label systems[10][9][11][18] of prior studies mainly lay emphasis on a limited scope of charge-oriented elements. The current element schema, especially those in the Chinese contexts, are far from enough. For example, victim number in highly likely a salient predictor in crime type and sentencing in Chinese criminal trials, yet we did not find any prior study in Chinese contexts that incorporate this label in their label system. Besides, existing studies predominantly focus on

*Equal contribution. Listing order is random.

[†]Corresponding

¹<https://wenshu.court.gov.cn>

²<https://home.pkulaw.com/>

the legally prescribed factors in sentencing, overlooking extra-legal elements. However, a wealth of empirical research suggests that these elements, such as the defendant’s and victim’s age, gender, race/ethnicity, etc., may significantly influence trial and sentencing outcomes [18][2][24][1][23]. The absence of these factors in the label system may compromise the performance of downstream tasks.

(2) **Lack of Domain Focus.** An overwhelming majority of existing datasets[12][16][26] for element extraction mainly focus on the element or event extraction in the general domain. However, such datasets may not be well suited to applications in the legal domain. For example, Recidivist (*Leifan* in Chinese) and Previous Criminal Record (*Qianke* in Chinese) are closely-connected yet distinct legal concepts in Chinese criminal law, which could be difficult to distinguish without adequate legal knowledge. Furthermore, various court participants may present different interpretations and perspectives on the same legal elements, such as whether the defendant voluntarily surrendered, confessed, or pled guilty. This diversity of opinion can cause confusion without professional legal annotation. Therefore, existing datasets from general domains are hardly applicable for comprehensive analysis and tasks based on legal texts owing to their lack of adequate understanding of legal knowledge and contexts.

(3) **Inadequate Few-Shot Charge Coverage.** Existing datasets[14] predominantly focus on high-frequency charges and often underperform when dealing with less frequent charges due to the limited number of cases. Moreover, charges that typically manifest with similar descriptions, such as the crime of forcible seizure and the crime of robbery, can be challenging to differentiate, especially with limited data.

(4) **Limited Application.** The previous datasets[30] in the legal domain do not consider the relations of elements, hindering the comprehension of judicial documents as they frequently present complex relationships. For example, in a document involving multiple defendants, the extracted labels of one defendant may not be applicable to another defendant, and thus, these labels should be linked to their corresponding defendant. The oversight of relational context frequently appear in judicial practice may significantly limit the performance of downstream tasks based on these datasets and their practical utility, including their usage in empirical analysis and effective LegalAI applications in real-world court settings.

To provide a solid foundation for legal element extraction, LEEC alleviates the above limitation in the following way:

(1) **Extensive Label System.** Our team of legal experts not only extended the coverage of legally prescribed factors that may significantly impact Chinese criminal trials and sentencing, but also actively drew upon a comprehensive collection of Chinese empirical studies published in Chinese core journals, as well as important legal papers in Chinese contexts published internationally. Based on the theory, research design, and findings of these empirical studies, we systematically compiled extra-legal key labels regarded by these studies as having substantial impact in Chinese judicial practice. In this way, we are able to constructed an extended and comprehensive label system in the legal domain.

(2) **Large Scale.** LEEC is annotated based on the publicly available cases of both LEVEN and LeCaRD, with a total of 15,831 cases. Therefore, the high coverage of cases could largely alleviate the

problem of limited number of cases in few-shot charges, leading to an increased ability of meeting needs in real-world court settings. Besides, the annotation of LEEC could be combined with the previous annotation from LEVEN and LeCaRD, providing more comprehensive information to facilitate the analysis of judicial documents.

(3) **Broader Application.** The knowledge graph we developed for our annotation system encapsulates significant relationships among various elements. For instance, as it is frequent for a single Chinese judicial document to involve multiple defendants, crimes, and victims, our team of legal experts has effectively linked defendant and victim characteristics to their respective individuals and affiliated crime characteristics to the corresponding offenses. This integration of crucial interrelations among labels enhances performance in various downstream applications, including the prediction of a specific defendant’s crime and sentencing, and also expands LEEC’s applicability in real-world court settings and future empirical research.

To validate the quality and applicability of LEEC, we implement various SOTA models in the field of document-level event extraction and evaluate them on our dataset, which shows that the elements of LEEC could be extracted with relatively high accuracy by these models.

2 DATA ANALYSIS

2.1 Corpus and Preprocessing

Our selection of cases draws from the publicly available LEVEN[31] and LeCaRD[15] datasets. These cases undergo a preprocessing phase wherein we extract the full text of each case, as well as their respective case number, crime, and year of judgement using automated algorithm, to facilitate the subsequent manual annotation. This complete dataset is comprised of 17,352 cases. After deduplication, the number of total unique cases is 17,231, encompassing 10,805 cases from LeCaRD, and 6,426 cases from LEVEN. All documents within this collection represent criminal decisions delivered in the span of the past two decades. Among the complete dataset, we preserve 1400 judicial document as non-published test dataset for future evaluation, and publish the remaining annotated data from 15,831 judicial documents. We compare LEEC with two types of datasets: (1) General-domain ED datasets. Compared with ACE2005[4] and MAVEN [27], LEEC is an element extraction dataset in the field of Chinese criminal law, with a label system designed for legal texts. The judicial documents of LEEC are also annotated by law school students with adequate understanding of legal knowledge and concepts. (2) Legal-domain datasets. Compared with LEVEN and LeCaRD, our dataset offers a comprehensive expansion of the label system with a finer level of granularity, encapsulating both legal and extra-legal labels. Furthermore, LEEC encompasses independent annotations of distinct entities such as victims, defendants, and crimes. This methodological approach considerably enhances the dataset’s precision and applicability, thereby contributing significantly to downstream applications and empirical research.

2.2 Data Distribution

Unlike previous datasets, LEEC incorporates information on multiple victims, defendants, and causes. It should be noted that a minor fraction of the annotated victim number contains missing values. This is attributable to instances where the precise number of victims cannot be ascertained due to the insufficient information provided within the judicial document. For a detailed explanation concerning the occurrence of missing values within the annotation, please see the Annotation section. Our dataset reveals the presence of multiple defendants in 34% of cases, multiple victims in 19% of all cases, and 40% of cases that contain at least one victim, and multiple crimes in 43% of cases. This underscores the necessity and effectiveness of introducing a sophisticated, domain-specific label system to handle such complexity. The distribution of cases is displayed in Table 1.

Defendant number	1	2	3	4	5	>5
Case distribution	10465	2262	1182	644	406	872
Victim number	0	1	2	3	4	>4
Case distribution	8140	4614	954	567	329	1226
Crime number	1	2	3	4	5	>5
Case distribution	9066	2920	1312	747	463	1323

Table 1: Case distribution on multi-defendants, victims, and crimes.

3 LABEL SYSTEM

This section encompasses the compilation of the extensive label system, as well as the establishment of the crucial relationships among those labels.

3.1 Label Compilation

Our team of legal experts, led by professors in law, incorporated a wide range of legal and extra-legal elements to build a comprehensive and extended knowledge graph covering key elements within the Chinese legal domain. First, our team of legal experts compiled the crucial legal circumstances and factors stipulated by Chinese criminal law and legal interpretations, such as whether the defendant confessed, pled guilty, voluntarily surrendered, conducted justifiable defense, etc.. Furthermore, it has long been underscored by researchers that there is to cope with the distinction between “law in books” and “law in action”, underscoring that the complexity of the actual application of law and its potential divergence from the written text[17]. It has widely been revealed that extra-legal factors may significantly impact the application of law, including judicial decisions, in practice. Therefore, we utilized elements and theories developed and validated by empirical legal research to comprehensively capture the important factors in Chinese criminal trials.

Specifically, our team of legal experts systematically compiled 178 quantitative legal studies from 2018 to 2022 published across 22 journals in Chinese listed in the China Legal Science Citation Index (CLSCI), an index curated by the Law Institute of China Law Society, providing the list of core legal journals in China. The majority of these studies investigated the impact of various legal and extra-legal factors on sentencing based on publicized judicial documents.

As such, the labels and theories used in these studies serve as a valuable source of factors that may significantly influence judicial decisions, thereby facilitating downstream tasks. In addition, we drew upon a wide range of empirical legal studies published in SSCI journals, especially those investigating sentencing factors in Chinese contexts. Our team meticulously collected the core theories and labels used in these studies and incorporated them into our legal system.

For instance, the Group Threat Theory suggests that when majority groups feel threatened by minority populations, criminal justice systems may treat racial or ethnic minorities adversely[25]. This theory has also been validated and developed in the Chinese context by prior empirical research, which found that minorities perceived as “problem minorities” that might disrupt public order may face discrimination in Chinese criminal cases[7]. Therefore, we included the ethnic status of offenders in our knowledge graph. Moreover, the Focal Concern Theory highlights the strategic function of judges in contexts of managerial uncertainty and constrained knowledge. It identifies three crucial factors influencing sentencing decisions: the defendant’s culpability, the risk posed to the community, and pragmatic considerations such as the court’s workload[21]. Research in Chinese contexts has shown that, in line with the Focal Concern Theory, defendant’s being a rural-to-urban migrant – measured by his or her registered permanent residence (*Hukou*) – significantly impacts sentencing outcomes[8]. As a result, we also included the registered permanent residence of defendants as an important element in our knowledge graph. Meanwhile, it should be noted that some elements could be easily and accurately extracted through identifying keywords or regular expression matching in Chinese criminal verdicts, and thus, do not require manual annotation, such as court name, judge name, case title, year of judgement, etc.. Therefore, these elements are not annotated or included in our knowledge graph and label system. Following this scheme, we effectively constructed an extended, multi-level knowledge graph to cover 159 important elements – both legal and extra-legal – in Chinese criminal sentencing. The elements in the knowledge graph are divided into four main categories: defendant characteristics, victim characteristics, case characteristics, and crime characteristics.

3.2 Relation Construction

We integrated the relations among elements into the knowledge graph we constructed, recognizing their significant influence on judicial decisions. This integration is particularly important in verdicts involving multiple crimes, offenders, victims, or defenders. To illustrate, consider a verdict with several defendants: the circumstances and characteristics of one defendant might differ from those of the others. Therefore, all characteristics pertaining to a victim or defendant are linked directly to the relevant individual. Besides, considering a defendant may have up to two defenders in Chinese criminal trials, the defender characteristics are also connected to the individual defender of a specific defendant. Furthermore, all characteristics of a crime are associated with the specific crime committed by a particular defendant. This is imperative because each defendant could be sentenced for multiple crimes in Chinese judicial documents. However, the circumstances of a specific crime may not necessarily apply to another. It is also noteworthy that in

Chinese criminal cases where an individual defendant committed multiple crimes, the court typically adjudicates a sentence for each individual crime, followed by an overall aggregated sentence. This final sentence, which is usually subject to a certain degree of the judge’s discretion, may not necessarily align with the sum of the individual sentences. Consequently, in our knowledge graph, we deliberately included both the sentencing elements, linked to each distinct crime of a specific defendant, and the final, aggregated sentence, linked to each defendant. The elements within the knowledge graph is depicted in Figure 1.

4 ANNOTATION

The annotation of LEEC requires the annotators to find and determine the element mentions, trigger words, values of each elements from the documents. Specifically, the annotation is conducted manually by a team of graduates and undergraduates majoring in law, trained and led by professors in law. All of them are interviewed before joining the team to ensure their ability to comprehend Chinese legal concepts and knowledge, and practiced for several hours before formal annotating. We have compiled a comprehensive 155-page annotation guideline in Chinese to assist our annotators. It provides an in-depth understanding of each element and their respective annotation methods. The guideline includes definitions, potential values for each element, common locations within the judicial documents where the element frequently appears, detailed rules for annotation, and real-world examples of document annotation for the elements, etc.. The examples of the annotation guideline are provided in detail in Appendix A.

During annotation, we adopted a two-stage process. In the first stage, we performed a fine annotation of 4182 documents randomly selected from the public datasets of LeCaRD and LEVEN. We annotated the element mentions, trigger words, and the values of each element, which are the basis for evaluating several baselines of the document-level event extraction task to validate the quality and applicability of the annotations for extraction tasks. The second stage is the extended annotation of the values of elements on the remaining public datasets of LeCaRD and LEVEN, which covers 11624 documents, to assist the verification of the results of element extraction tasks. Besides, as the label system of LEEC is largely originated from prior empirical studies, this extended dataset could also contribute to the replication and exploration of empirical research. In both stages, we performed double annotations on a portion of the work of each annotator to check the consistency and quality of the annotation. Datasets from both stages are publicly available on <https://github.com/THUlawtech/LEEC>.

During the annotation process, we observed that a minor fraction of judicial documents contained an unusually high number of defendants or victims, in some cases reaching into the hundreds. These documents were predominantly associated with corporate crimes committed for financial gains. In response to this situation, we implemented an upper limit, treating cases with more than seven defendants or victims as if they had exactly seven, and only annotated the first seven defendants and victims.

It is also noteworthy that a number of elements may not always be explicitly mentioned in judicial documents. For instance, the offender gender, while frequently disclosed, is not always explicitly

stated, as discretion is commonly exercised in such circumstances. Similarly, whether an offender has received forgiveness from victims or their close relatives is sometimes clearly affirmed or negated, yet such information may be not mentioned in many judicial documents. As a general rule, we classify these non-mention instances as missing values. However, an exception exists when our team of legal experts determines that, in almost all cases in Chinese judicial practice, a certain element (of binary nature) is specified in the judicial document when it holds a value of 1. Thus, its absence from a judicial document suggests that the element holds a value of 0. An example is the COUNTERCLAIM element, which denotes whether the defendant has lodged a counterclaim against the private prosecutor or the victim in the incidental civil action portion of the criminal case. In practice, this element is typically documented if and only if a counterclaim has indeed been initiated. Hence, should the element be absent from the judicial document, the annotator would assign it a value of 0. The same principle applies to elements concerning the sentencing type of offenders. If the offender does not receive a specific type of punishment, the judicial document would not mention the corresponding element in the sentencing section. Therefore, such non-mention instances are also assigned a value of 0 during annotation. Tables 5 through 15 contain special notes for elements wherein non-mention does not equate to missing value.

We measure the data quality by Kappa, with a value of 0.71. This value demonstrates that the manual annotation of LEEC is conducted with high quality, contributing to the development of legal element extraction and the analysis of legal cases.

5 DOCUMENT-LEVEL EVENT EXTRACTION

This section delves into the specifics of the experiment conducted using the LEEC dataset. It covers the experiment’s background and settings, the suitable baselines and metrics, as well as the results and associated discussions.

5.1 Experiment Settings

For the DEE task, we selected some representative labels in LEEC labeling system to extract important event information of the defendant, and the parameters in the event table are shown in Table 2. Since most legal cases do not typically demonstrate a direct correspondence relation between defendants and victims when there are multiple defendants and victims involved, we uniformly assign the name of the first victim appearing in the document as the victim’s name. The maximum number of defendants to be drawn from a judgment document is 7. After the above filtering steps, the LEEC-DEE dataset size is 4156, with 524 annotated documents from LEVEN and 3632 annotated documents from LeCaRD. We split the dataset into the training, validation and test sets at a proximate ratio of 8:1:1. We use the same vocabulary as [32] and randomly initialize all the embeddings where $dh=768$ and $dl=32$. We employ the Adam optimizer with the learning rate $2e-5$ and the batchsize is 64. All models are trained for 100 epochs and the checkpoints with the best F1 scores on the dev set are selected for evaluation on the test set.

Defendant	Role Type	Corresponding label
Demographic Characteristics	Name	Defendant_name
	Gender	Defendant_gender
	Birth	Defendant_birth
	Nation	Defendant_nationality
	Place	Defendant_birthplace
Aggregated Sentencing	Control	Probation_aggregated
	ControlTime	Probation_term_aggregated
	Detention	Limited_incarceration_aggregated
	DetentionTime	Limited_incarceration_term_aggregated
	Imprisonment	Fixed_imprison_aggregated
	ImprisonmentTime	Fixed_imprison_term_aggregated
	PoliticalRights	Political_deprivation_aggregated
	PoliticalRightsTime	Political_deprivation_term_aggregated
	Fine	Fine_aggregated
	FineNum	Fine_amount_aggregated
	PartofProperty	Partial_property_confiscation_aggregated
	PartofPropertyNum	Partial_confiscated_amount_aggregated
	AllProperty	Total_property_confiscation_aggregated
	AllPropertyNum	Total_confiscated_amount_aggregated
	EcoCompensation	Loss_compensation_aggregated
	EcoCompensationNum	Compensation_amount_aggregated
Victim information	VictimName	Victim_name

Table 2: Roles in the defendant event table and their corresponding labels in the LEEC label system.

5.2 Baselines and Metrics

Baselines. We introduce the following typical models as baselines: (1) DCFEE[29] is the first model introduced Distance Supervision (DS) to solve DEE task. there are two variants included: DCFEE-O only extracts one event record from one document while DCFEE-M tries to extract multiple possible event records. (2) Doc2EDAG[32] is an end-to-end DEE model that constructs event records in an autoregressive way by generating entity-based Directed Acyclic Graphs (DAGs). (3) GreedyDec is a baseline proposed in Doc2EDAG[32] which fills one event table greedily. (4) PTPCG[33] is a lightweight model for end-to-end document-level event extraction based on pruned complete graphs with pseudo triggers.

Metrics. Following the same evaluation setting in the previous studies[32][33]. For each prediction record, we select a golden record by matching records with the same event type and the most shared arguments, and calculate the F1 score by comparing the parameters between them.

5.3 Results

Table 3 shows the experimental results on LEEC-DEE dataset. Table 4 presents the the epochs where the baselines perform optimally. From experiments we have the following observations: 1) The ability of most models is consistent with their performance on previous DEE data-sets[33]; 2) Some baselines cannot converge well on datasets, such as Doc2EDAG[32] and similar structured GreedyDec. One reason is that compared to the previously evaluated data in the general or financial fields, the text of legal documents is longer and the arguments are more dispersed, which is not conducive to the Doc2EDA-G[32] structure using a sequential path extension method for reasoning. Experiments show that LEEC-DEE dataset poses challenges to DEE models, indicating that DEE in legal domain is an open issue.

Model	Precision	Recall	F1 score
DCFEE-O	67.37%	77.90%	72.25%
DCFEE-M	67.03%	72.85%	69.82%
Greedy-Dec	79.63%	59.25%	67.94%
Doc2EDAG	26.61%	72.39%	38.91%

Table 3: Overall performance on Document-level Event Extraction.

Model	Train Epoch	Best Epoch
DCFEE-S	100	24
DCFEE-M	100	15
Greedy-Dec	100	6
Doc2EDAG	100	8
PTPCG	100	88

Table 4: The train epoch and best epoch in which the models achieve the highest micro F1 score on the dev set.

6 DISCUSSION AND CONCLUSION

In this study, we introduce LEEC, a uniquely tailored dataset designed for the extraction of legal elements within the Chinese criminal law system. Our dataset stands out from existing element extraction datasets as it is enriched with an expansive legal domain label system, meticulously curated by our team of legal experts. This system integrates crucial legal knowledge drawn from Chinese law, empirical legal studies, and an understanding of the legal contexts derived from judicial practices. Each of the 15,831 cases in the dataset has been annotated by law school students. Experimental results underline the challenges associated with multi-label prediction, signifying an area of focus for future research.

It is pertinent to note that the knowledge graph and label system developed in this study exhibit a notably low level of granularity. This specificity was designed to support valuable downstream applications and empirical legal research. Yet, it is of paramount importance that users of LEEC exercise due caution. We vehemently oppose to the use of LEEC for any purposes that could lead to discrimination or violate the principle of the rule of law, whether within or outside the courtroom. The personal information included in the published judicial documents was collected and processed in strict compliance with Chinese law. Any future utilization of LEEC must also adhere to applicable laws and commit to responsible, ethical handling of this data to prevent misuse and uphold the principle of privacy. As it is crucial that all users understand their role in maintaining ethical standards and protecting personal information, we urge each user to consider the potential implications of their work and to use the dataset of this study responsibly.

This study has several limitations that we hope will be addressed by future research: 1) We selected only 22 elements for the DEE task as the corresponding text of these elements could be clearly and directly extracted from the document, such as the defendant’s name. However, some elements within the system cannot always be directly extracted, such as the JOINT_CRIME element, which can be expressed in a flexible manner within judicial documents. Consequently, the annotation of this element necessitates the specification of complex rules and, occasionally, a certain degree of

reading comprehension. Developing models to extract such labels may present a significant challenge; 2) The study inherently deals with sparse matrix for some elements that appear infrequently in the judicial documents, such as the LEGAL_AID element, which may pose challenges in the DEE task. The sparsity of data can lead to computational difficulties and poor model performance. Future research should consider strategies for handling and interpreting sparse data. 3) Based on estimations, approximately 75% of all judicial verdicts in China were ultimately disclosed for cases not processed through criminal mediation in recent years [22]. This suggests that our dataset may inevitably contain selection bias. Therefore, the distribution and correlations of labels in our dataset may not fully represent those in actual Chinese courts. Future studies should consider this potential bias when interpreting findings based on this dataset.

ACKNOWLEDGEMENTS

We extend our sincere gratitude to Huaiyu Hu and Senyu Li from the Law School at Tsinghua University for their invaluable legal assistance. Meanwhile, we wish to express our deep appreciation to Zexia Yang and Ziheng Xie, also from the Law School at Tsinghua University, for their help and advice regarding data processing and knowledge graph construction. This work is supported by the National Key Research and Development Program of China (No. 2022YFC3301504).

REFERENCES

- [1] Andong Chen, Feng Yao, Xinyan Zhao, Yating Zhang, Changlong Sun, Yun Liu, and Weixing Shen. 2023. Equals: a real-world dataset for legal question answering via reading chinese laws. In *Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law*, 71–80.
- [2] Jill K Doerner and Stephen Demuth. 2010. The independent and joint effects of race/ethnicity, gender, and age on sentencing outcomes in us federal courts. *Justice Quarterly*, 27, 1, 1–27.
- [3] Yi Feng, Chuanyi Li, and Vincent Ng. 2022. Legal judgment prediction via event extraction with constraints. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics, Dublin, Ireland, (May 2022), 648–664. doi: 10.18653/v1/2022.acl-long.48.
- [4] Ralph Grishman, David Westbrook, and Adam Meyers. 2005. Nyu’s english ace 2005 system description. *ACE*, 5, 2.
- [5] Kaihao Guo, Tianpei Jiang, and Haipeng Zhang. 2020. Knowledge graph enhanced event extraction in financial documents. In *2020 IEEE International Conference on Big Data (Big Data)*. IEEE, 1322–1329.
- [6] Frederik Hogenboom, Flavius Frasinca, Uzay Kaymak, and Franciska De Jong. 2011. An overview of event extraction from text. *DeRIVE@ ISWC*, 48–57.
- [7] Yue Hou and Rory Truex. 2020. Ethnic discrimination and authoritarian rule: an analysis of criminal sentencing in china. Available at SSRN 3481448.
- [8] Jize Jiang and Kai Kuang. 2018. Hukou status and sentencing in the wake of internal migration: the penalty effect of being rural-to-urban migrants in china. *Law & Policy*, 40, 2, 196–215.
- [9] Haitao Li, Qingyao Ai, Jia Chen, Qian Dong, Yueyue Wu, Yiqun Liu, Chong Chen, and Qi Tian. 2023. Sailer: structure-aware pre-trained language model for legal case retrieval. (2023). arXiv: 2304.11370 [cs. IR].
- [10] Haitao Li, Weihang Su, Changyue Wang, Yueyue Wu, Qingyao Ai, and Yiqun Liu. 2023. Thuir@coliee 2023: incorporating structural knowledge into pre-trained language models for legal case retrieval. (2023). arXiv: 2305.06812 [cs. IR].
- [11] Haitao Li, Changyue Wang, Weihang Su, Yueyue Wu, Qingyao Ai, and Yiqun Liu. 2023. Thuir@coliee 2023: more parameters and legal knowledge for legal case entailment. (2023). arXiv: 2305.06817 [cs. CL].
- [12] Xinyu Li, Fayuan Li, Lu Pan, Yuguang Chen, Weihua Peng, Quan Wang, Yajuan Lyu, and Yong Zhu. 2020. Duee: a large-scale dataset for chinese event extraction in real-world scenarios. In *Natural Language Processing and Chinese Computing: 9th CCF International Conference, NLPCC 2020, Zhengzhou, China, October 14–18, 2020, Proceedings, Part II 9*. Springer, 534–545.
- [13] Shasha Liao and Ralph Grishman. 2010. Using document level cross-event inference to improve event extraction. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, 789–797.
- [14] Bulou Liu, Yiran Hu, Yueyue Wu, Yiqun Liu, Fan Zhang, Chenliang Li, Min Zhang, Shaoping Ma, and Weixing Shen. 2023. Investigating conversational agent action in legal case retrieval. In *European Conference on Information Retrieval*. Springer, 622–635.
- [15] Yixiao Ma, Yunqiu Shao, Yueyue Wu, Yiqun Liu, Ruizhe Zhang, Min Zhang, and Shaoping Ma. 2021. Lecard: A legal case retrieval dataset for chinese law system. In *SIGIR ’21: The 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, Canada, July 11–15, 2021*. Fernando Diaz, Chirag Shah, Torsten Suel, Pablo Castells, Rosie Jones, and Tetsuya Sakai, (Eds.) ACM, 2342–2348. doi: 10.1145/3404835.3463250.
- [16] Kiem-Hieu Nguyen, Xavier Tannier, Olivier Ferret, and Romaric Besançon. 2016. A dataset for open event extraction in english. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC’16)*, 1939–1943.
- [17] Roscoe Pound. 1910. Law in books and law in action. *American Law Review*, 44, 12.
- [18] Tara N Richards, Wesley G Jennings, M Dwayne Smith, Christine S Sellers, Sondra J Fogel, and Beth Bjerregaard. 2016. Explaining the “female victim effect” in capital punishment: an examination of victim sex-specific models of juror sentence decision-making. *Crime & Delinquency*, 62, 7, 875–898.
- [19] Shirong Shen, Guilin Qi, Zhen Li, Sheng Bi, and Lusheng Wang. 2020. Hierarchical Chinese legal event extraction via pedal attention mechanism. In *Proceedings of the 28th International Conference on Computational Linguistics*. International Committee on Computational Linguistics, Barcelona, Spain (Online), (Dec. 2020), 100–113. doi: 10.18653/v1/2020.coling-main.9.
- [20] G Sierra et al. 2018. Event extraction from legal documents in spanish. In *1st Workshop on Language Resources and Technologies for the Legal Knowledge Graph*, 36.
- [21] Darrell Steffensmeier and Stephen Demuth. 2000. Ethnicity and sentencing outcomes in us federal courts: who is punished more harshly? *American sociological review*, 705–729.
- [22] Yingmao Tang and John Zhuang Liu. 2019. Mass publicity of chinese court decisions. *China Review*, 19, 2, 15–40.
- [23] Vu Tran, Minh Le Nguyen, and Ken Satoh. 2019. Building legal case retrieval systems with lexical matching and summarization using a pre-trained phrase scoring model. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*, 275–282.
- [24] Jeffery T Ulmer. 2012. Recent developments and new directions in sentencing research. *Justice Quarterly*, 29, 1, 1–40.
- [25] Jeffery T Ulmer and Brian Johnson. 2004. Sentencing in context: a multilevel analysis. *Criminology*, 42, 1, 137–178.
- [26] Amir Pouran Ben Veyseh, Javid Ebrahimi, Franck Dernoncourt, and Thien Huu Nguyen. 2022. Mee: a novel multilingual event extraction dataset. *arXiv preprint arXiv:2211.05955*.
- [27] Xiaozhi Wang et al. 2020. MAVEN: A massive general domain event detection dataset. In *Proceedings of EMNLP*, 1652–1671. doi: 10.18653/v1/2020.emnlp-main.129.
- [28] Chaojun Xiao et al. 2019. CAIL2019-SCM: A dataset of similar case matching in legal domain. *CoRR*, abs/1911.08962. <https://arxiv.org/abs/1911.08962> arXiv: 1911.08962.
- [29] Hang Yang, Yubo Chen, Kang Liu, Yang Xiao, and Jun Zhao. 2018. Dcfee: a document-level chinese financial event extraction system based on automatically labeled training data. In *Proceedings of ACL 2018, System Demonstrations*, 50–55.
- [30] Feng Yao, Jingyuan Zhang, Yating Zhang, Xiaozhong Liu, Changlong Sun, Yun Liu, and Weixing Shen. 2023. Unsupervised legal evidence retrieval via contrastive learning with approximate aggregated positive. In *Proceedings of the AAAI Conference on Artificial Intelligence* number 4. Vol. 37, 4783–4791.
- [31] Feng Yao et al. 2022. LEVEN: A large-scale chinese legal event detection dataset. In *Findings of ACL*, 183–201. doi: 10.18653/v1/2022.findings-acl.17.
- [32] Shun Zheng, Wei Cao, Wei Xu, and Jiang Bian. 2019. Doc2edag: an end-to-end document-level framework for chinese financial event extraction. *arXiv preprint arXiv:1904.07535*.
- [33] Tong Zhu, Xiaoye Qu, Wenliang Chen, Zhefeng Wang, Baoxing Huai, Nicholas Jing Yuan, and Min Zhang. 2021. Efficient document-level event extraction via pseudo-trigger-aware pruned complete graph. *arXiv preprint arXiv:2112.06013*.

A EXAMPLES FROM THE ANNOTATION GUIDELINE

A.1 Joint_crime

1. Label Meaning

A JOINT_CRIME element refers to whether the court determined that an intentional crime was committed by two or more persons or units in collaboration.

2. Potential Value of Element

1 = Yes; 0 = No

3. Annotation Rules

- (1) This element generally appears in the fact description section (typically after the statement of “The court finds”) or the reasoning section (typically after the statement of “The court considers”).
- (2) This element cannot be determined solely based on the number of defendants that appear in the judicial document. Having multiple defendants does not necessarily constitute a joint crime; having only one defendant does not necessarily mean there is no joint crime.
- (3) If the judicial document explicitly states that “the behavior is a joint crime”, then it is a joint crime. If there are other defendants mentioned in the same criminal act, but handled separately, it also counts as a joint crime.
- (4) If terms like “in league with others” or “conspiring with others” appear, it is generally considered a joint crime. Any ambiguities or uncertainties should be reported to us.
- (5) If the document does not explicitly state that the crime is a joint crime, but contains phrases such as the defendant “is an accomplice”, “is a principal offender”, it is considered a joint crime.
- (6) If the joint crime status is inconsistent across different charges, each crime should be annotated separately.
- (7) If it can be clearly determined from the judicial document that there is only one perpetrator, it can be determined that it is not a joint crime.
- (8) If a crime contains multiple criminal acts, but only some of the criminal facts are jointly committed, the rest are not, it may still be considered a joint crime.
- (9) In corporate crimes, the entity and its directly responsible personnel may be both considered guilty by the court, but this does not constitute a joint crime.

4. Annotation Example

Original Document Text

Case number: (2011) Yong Zhen Xing Chu Zi No.49

From November to December 1994, the defendant FANG Xingdu and FANG Jinqi (already sentenced) **conspired in advance**, using the convenience of FANG Xingdu’s position as a guard at the original Ningbo Heqiao Chemical Co., Ltd. (now Ningbo Xinqiao Chemical Co., Ltd.), responsible for the receipt of styrene raw materials, when FANG Jinqi drove Ningbo Chemical Hazardous Goods Transport Company’s chemical tank truck to transport styrene raw materials from Zhenhai Port Area to Ningbo Heqiao Chemical Co., Ltd... The court believes that the defendant FANG Xingdu **conspired with others**...

Annotation Result

1

Reason for Annotation

“Conspired in advance” indicates that the two defendants conspired

with each other in advance; “conspired with others” is also a typical expression of joint crime. Therefore, even if the judgment document does not explicitly state that the defendant committed a “joint crime”, it can also be marked 1.

A.2 Forgiveness

1. Label Meaning

A FORGIVENESS element refers to whether the court determined that the defendant had obtained forgiveness from victims or their close relatives.

2. Potential Value of Element

1 = Yes; 0 = No

3. Annotation Rules

- (1) This element generally appears in the fact description section (typically after the statement of “The court finds”) or the reasoning section (typically after the statement of “The court considers”).
- (2) For this element to be assigned a value of 1, the document usually contains expressions such as “obtained the forgiveness of the victim’s relatives” or “obtained the forgiveness of the victim”. The forgiveness here includes the victims and their close relatives.
- (3) The element should be annotated for each defendant who appears in the judicial document, respectively.
- (4) If there are multiple defendants in the case, and one of them obtains forgiveness, it does not mean that the victim also forgives other defendants.
- (5) If there are multiple victims in the case, and only some of the victims forgave the defendant, this still constitutes the circumstance of obtaining forgiveness, and thus, the element should be assigned a value of 1.

4. Annotation Example

Original Document Text

Case Number: (2017) Yu 1381 Criminal First Instance 426

The defendants, WANG Congwen and HUANG Jinlian, confessed their crimes truthfully, and thus, could receive lighter punishments. They compensated the economic losses of the victim’s close relatives, and **obtained the forgiveness of the victim’s close relatives**. Therefore, they can be punished lightly at discretion. The defendant HUANG Jinlian has shown remorse and has no danger of reoffending. A suspended sentence has no significant adverse impact on the community where she lives...

Annotation Result

1

Reason for Annotation

The document clearly states that the defendants obtained the forgiveness of the victim’s relatives.

B ELEMENT SCHEMA AND DESCRIPTION

To facilitate the understanding of our label system and each element within it, thereby promoting future application and research, we provide the multi-level label system in Figure 1. Additionally, Tables 5 to 15 provide detailed descriptions of each element, including element name, explanation, and value type.

LEEC: A Legal Element Extraction Dataset with an Extensive Domain-Specific Label System

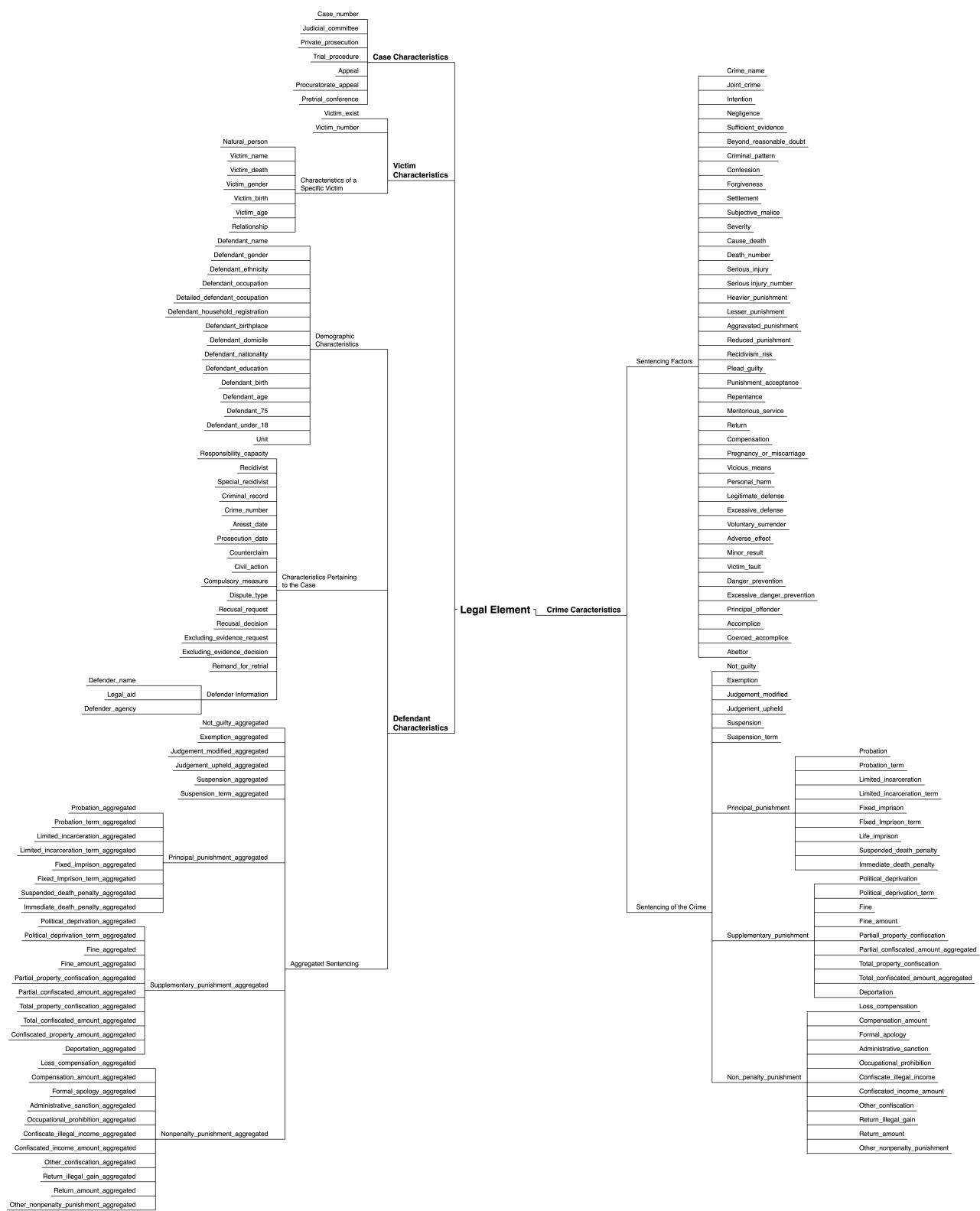


Figure 1: The Detailed Element Schema of LEEC.

Element Name	Explanation	Value Type
Case Characteristics		
Case_number	A CASE_NUMBER element refers to the ID assigned by the court that uniquely identifies each judicial documents.	Extracted from the Specific Content of Judicial Documents
Judicial_committee	A JUDICIAL_COMMITTEE element refers to whether the court submitted the case to the judicial committee for discussion.	1 = Yes, 0 = No
Private_prosecution	A PRIVATE_PROSECUTION element refers to whether the victim, the victim's legal representative, or close relative institute an action directly in a people's court in a case of private prosecution.	1 = Yes, 0 = No, Non-mention is treated as 0
Trial_procedure	A TRIAL_PROCEDURE element refers to the procedure applied to the trial of the case.	Summary, Formal, Fast-Track Sentencing, Transfer from Summary to Formal Procedure, Transfer from Fast-Track to Formal Procedure
Appeal	AN APPEAL element refers to whether the defendant appealed the case.	1 = Yes, 0 = No
Procuratorate_appeal	A PROCURATORATE_APPEAL element refers to whether the people's procuratorate files an appeal to the people's court at the higher level.	1 = Yes, 0 = No
Pretrial_conference	A PRETRIAL_CONFERENCE element refers to whether the court determined that a pretrial conference for a case should be held.	1 = Yes, 0 = No
Victim Characteristics (I)		
Victim_exist	A VICTIM_EXIST element refers to there was a victim whose legal rights and interests have been violated by a criminal act in a criminal case according to the court.	1 = Yes, 0 = No
Victim_number	A VICTIM_NUMBER element refers to the number of victims according to the information provided by the court. This element is annotated only when elements VICTIM_EXIST is assigned a value of 1.	Numeric Value Calculated Based on the Content of Judicial Documents
Natural_person	A NATURAL_PERSON element refers to whether the victim is a natural person in the biological sense. This element is annotated only when elements VICTIM_EXIST is assigned a value of 1.	1 = Yes, 0 = No
Victim_name	A VICTIM_NAME element refers to the name of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Victim_death	A VICTIM_DEATH element refers to whether a specific victim was dead. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	1 = Yes, 0 = No
Victim_gender	A VICTIM_GENDER element refers to the gender of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Victim_birth	A VICTIM_BIRTH element refers to the date of birth of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Victim_age	A VICTIM_AGE element refers to the age of the victim according to the court. This element is annotated only when elements VICTIM_EXIST and NATURAL_PERSON are both assigned a value of 1.	Extracted from the Specific Content of Judicial Documents

Table 5: List of detailed element information (I).

Element Name	Explanation	Value Type
Victim Characteristics (II)		
Relationship	A RELATIONSHIP element refers to the relationship between the defendant and the victim based on the content of judicial documents.	Relationship between Non-natural Persons, Marital Relationship, Close Relatives, Other Relatives, Acquainted, Unknown
Defendant Characteristics (I)		
Defendant_name	A DEFENDANT_NAME element refers to the name of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_gender	A DEFENDANT_GENDER element refers to the gender of the defendant.	1 = Male; 0 = Female
Defendant_ethnicity	A DEFENDANT_ETHNICITY element refers to the ethnicity of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_occupation	A DEFENDANT_OCCUPATION element refers to the occupation of the defendant categorized into four types.	Extracted from the Specific Content of Judicial Documents
Detailed_defendant_occupation	A DETAILED_DEFENDANT_OCCUPATION element refers to the detailed occupation of the defendant as explicitly stated in the judicial document.	Extracted from the Specific Content of Judicial Documents
Defendant_household_registration	A DEFENDANT_HOUSEHOLD_REGISTRATION element refers to the place of registered permanent residence of the defendant, also known as <i>Hukou</i> in Chinese.	Extracted from the Specific Content of Judicial Documents
Defendant_birthplace	A DEFENDANT_BIRTHPLACE element refers to the place of birth of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_domicile	A DEFENDANT_DOMICILE element refers to the address of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_nationality	A DEFENDANT_NATIONALITY element refers to the nationality of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_education	A DEFENDANT_EDUCATION element refers to the level of formal education or academic degree attained by the defendant.	Illiteracy, Under Secondary School, Secondary School, Primary School, Regular Senior High School, Secondary Vocational School, Tertiary Vocational School, Bachelor's Degree, Master's Degree, Doctor's Degree
Defendant_birth	A DEFENDANT_BIRTH element refers to the date of birth of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_age	A DEFENDANT_AGE element refers to the age of the defendant.	Extracted from the Specific Content of Judicial Documents
Defendant_75	A DEFENDANT_75 element refers to whether the court determined that the defendant is above the age of 75, which is a mitigating circumstance in Chinese criminal law, and thereby may be explicitly stated in the judicial document.	1 = Yes; 0 = No
Defendant_under_18	A DEFENDANT_UNDER_18 element refers to whether the court determined that the defendant is under the age of 18, which is a mitigating circumstance in Chinese criminal law, and thereby may be explicitly stated in the judicial document.	1 = Yes; 0 = No
Unit	A UNIT element refers to whether the defendant is a company, enterprise, institution, organization, or group.	1 = Yes; 0 = No
Responsibility_capacity	A RESPONSIBILITY_CAPACITY element refers to the court determined that the level of the defendant's ability to take responsibility for crimes.	Full Criminal Responsibility Capacity Criminal Responsibility Incapacity Relatively Criminal Responsibility Incapacity Partial Criminal Responsibility Capacity
Recidivist	A RECIDIVIST element refers to whether the court determined that the defendant was a recidivist.	1 = Yes; 0 = No

Table 6: List of detailed element information (II).

Element Name	Explanation	Value Type
Defendant Characteristics (II)		
Special_recidivist	A SPECIAL_RECIDIVIST element refers to whether the court determined that the defendant was a special recidivist as stipulated in the Chinese criminal law.	1 = Yes; 0 = No
Criminal_record	A CRIMINAL_RECORD element refers to whether the court determined that the defendant had a previous criminal record.	1 = Yes; 0 = No
Crime_number	A CRIME_NUMBER element refers to the total number of the crime name of a specific defendant.	Numeric Value Calculated Based on the Content of Judicial Documents
Arrest_date	An ARREST_DATE element refers to the date of the execution of arrest by criminal justice authorities.	Extracted from the Specific Content of Judicial Documents
Prosecution_date	A PROSECUTION_DATE element refers to the date of the initiation of a public prosecution by the people's procuratorate.	Extracted from the Specific Content of Judicial Documents
Counterclaim	A COUNTERCLAIM element refers to whether the defendant files a counterclaim against the private prosecutor in a private criminal prosecution case or the victim in the incidental civil action part of the criminal incidental civil case.	1 = Yes; 0 = No; Non-mention is treated as 0
Civil_action	An CIVIL_ACTION element refers to whether the court determined that the incidental civil action is instituted.	1 = Yes; 0 = No; Non-mention is treated as 0
Compulsory_measure	A COMPULSORY_MEASURE element refers to the methods to restrict a certain degree of personal freedom for criminal suspects and defendants by criminal justice authorities.	Custody/Forced Appearance/Granted Bail/Residential Confinement/Detention/Arrest
Dispute_type	A DISPUTE_TYPE element refers to the type of dispute involved in the case, including disputes with clearly identified victims, disputes between neighbors, family disputes, and disputes without clearly identified victims.	Family Disputes / Disputes among Neighbors / Other Disputes with Victims / Disputes without Victims
Recusal_request	A RECUSAL_REQUEST element refers to whether the parties requested for recusal.	1 = Yes; 0 = No
Recusal_decision	A RECUSAL_DECISION element refers to whether the court approved of a recusal request.	1 = Yes; 0 = No
Excluding_evidence_request	AN EXCLUDING_EVIDENCE_APPLICATION element refers to whether parties and their defenders or litigation representatives requested for excluding illegal evidence.	1 = Yes; 0 = No
Excluding_evidence_decision	AN EXCLUDING_EVIDENCE_DECISION element refers to whether the court approved of a request for excluding illegal evidence.	1 = Yes; 0 = No
Remand_for_retrial	A REMAND_FOR_RETRIAL element refers to whether the court determined that the case shall be remanded for retrial.	1 = Yes; 0 = No
Defender_name	A DEFENDER_NAME element refers to the name of the defender.	Extracted from the Specific Content of Judicial Documents
Legal_aid	A LEGAL_AID element refers to whether the defender was designated by the legal aid agency.	1 = Yes; 0 = No
Defender_agency	A DEFENDER_AGENCY element refers to the agency of the defender.	Extracted from the Specific Content of Judicial Documents
Not_guilty_aggregated	A NOT_GUILTY_AGGREGATED element refers to whether the court determined that the defendant was not guilty.	1 = Yes; 0 = No; Non-mention is treated as 0
Exemption_aggregated	An EXEMPTION_AGGREGATED element refers to whether the defendant is exempted from criminal punishment as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0

Table 7: List of detailed element information (III).

Element Name	Explanation	Value Type
Defendant Characteristics (III)		
Judgement_modified_aggregated	A JUDGEMENT_MODIFIED_AGGREGATED element refers to whether the court determined that full or partial revision of sentence is made in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Judgement_upheld_aggregated	A JUDGEMENT_UPHELD_AGGREGATED element refers to whether the court decided to uphold the previous judgment in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspension_aggregated	A SUSPENSION_AGGREGATED element refers to whether the court determined a suspension of sentence in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspension_term_aggregated	A SUSPENSION_TERM_AGGREGATED element refers to the probation period for suspension as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element SUSPENSION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Probation_aggregated	A PROBATION_AGGREGATED element refers to whether the defendant was subject to probation as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Probation_term_aggregated	A PROBATION_TERM_AGGREGATED element refers to the term of probation as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element PROBATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Limited_incarceration_aggregated	A LIMITED_INCARNATION_AGGREGATED element refers to whether the defendant was subject to limited incarceration as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Limited_incarceration_term_aggregated	A LIMITED_INCARCERATION_TERM_AGGREGATED element refers to the term of limited incarceration as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element LIMITED_INCARNATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Fixed_imprison_aggregated	A FIXED_IMPRISON_AGGREGATED element refers to whether the defendant was subject to fixed incarceration as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Fixed_imprison_term_aggregated	A FIXED_IMPRISON_TERM_AGGREGATED element refers to the term of fixed-term imprisonment as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element FIXED_IMPRISON_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Life_imprison_aggregated	A LIFE_IMPRISON_AGGREGATED element refers to whether the defendant was subject to life imprisonment as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspended_death_penalty_aggregated	A SUSPENDED_DEATH_PENALTY_AGGREGATED element refers to whether the defendant was subject to a suspended death penalty as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Immediate_death_penalty_aggregated	A IMMEDIATE_DEATH_PENALTY_AGGREGATED element refers to whether the defendant was subject to an immediate death penalty as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0

Table 8: List of detailed element information (IV).

Element Name	Explanation	Value Type
Defendant Characteristics (IV)		
Political_deprivation_aggregated	A POLITICAL_DEPRIVATION_AGGREGATED element refers to whether the defendant was subject to the deprivation of political rights as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Political_deprivation_term_aggregated	A POLITICAL_DEPRIVATION_TERM_AGGREGATED element refers to the term of deprivation of political rights as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element POLITICAL_DEPRIVATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Fine_aggregated	A FINE_AGGREGATED element refers to whether the defendant was fined as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Fine_amount_aggregated	A FINE_AMOUNT_AGGREGATED element refers to the amount of the fine as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element FINE_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Partial_property_confiscation_aggregated	A PARTIAL_PROPERTY_CONFISCATION_AGGREGATED element refers to whether the defendant was subject to confiscation of a part of his or her property as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Total_property_confiscation_aggregated	A TOTAL_PROPERTY_CONFISCATION_AGGREGATED element refers to whether the defendant was subject to confiscation of all of his or her property as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscated_property_amount_aggregated	A CONFISCATED_PROPERTY_AMOUNT_AGGREGATED element refers to the amount of confiscated property as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when either element PARTIAL_PROPERTY_CONFISCATION_AGGREGATED or element TOTAL_PROPERTY_CONFISCATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Deportation_aggregated	A DEPORTATION_AGGREGATED element refers to whether the defendant was subject to deportation as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Loss_compensation_aggregated	A LOSS_COMPENSATION_AGGREGATED element refers to whether the defendant was ordered to make compensation for the economic loss as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Compensation_amount_aggregated	A COMPENSATION_AMOUNT_AGGREGATED element refers to the amount of making compensation for the economic loss as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element LOSS_COMPENSATION_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Formal_apology_aggregated	A FORMAL_APOLOGY_AGGREGATED element refers to whether the defendant was ordered to make a statement of repentance or formal apology as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Administrative_sanction_aggregated	An ADMINISTRATIVE_SANCTION_AGGREGATED element refers to whether the defendant was subjected to administrative sanctions by the relevant department as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0

Table 9: List of detailed element information (V).

Element Name	Explanation	Value Type
Defendant Characteristics (V)		
Occupational_prohibition_aggregated	An OCCUPATIONAL_PROHIBITION_AGGREGATED element refers to whether the defendant was subjected to occupational prohibition as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscate_illegal_income_aggregated	A CONFISCATE_ILLEGAL_INCOME_AGGREGATED element refers to whether the defendant was subjected to the confiscation of illegal income as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscated_income_amount_aggregated	A CONFISCATED_INCOME_AMOUNT_AGGREGATED element refers to the amount of confiscated income as determined by the court in the aggregated sentencing of a defendant. This element is annotated only when element CONFISCATE_ILLEGAL_INCOME_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_confiscation_aggregated	An OTHER_CONFISCATION_AGGREGATED element refers to whether the defendant was subjected to the confiscation of other objects related to the crime as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0
Return_illegal_gain_aggregated	A RETURN_ILLEGAL_GAIN_AGGREGATED element refers to the order for returning illegal gains and compensations as determined by the court in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Return_amount_aggregated	A RETURN_AMOUNT_AGGREGATED element refers to the amount of returning illegal gains and compensations as ordered by the court in the aggregated sentencing of a defendant. This element is annotated only when element RETURN_ILLEGAL_GAIN_AGGREGATED is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_nonpenalty_punishment_aggregated	A OTHER_NONPENALTY_PUNISHMENT_AGGREGATED element refers to the order of other types of nonpenalty punishment as determined by the court in the aggregated sentencing of a defendant.	1 = Yes; 0 = No; Non-mention is treated as 0
Crime Characteristics (I)		
Crime_name	A CRIME_NAME element refers to the charges against the defendant.	Extracted from the Specific Content of Judicial Documents
Joint_crime	A JOINT_CRIME element refers to whether the court determined that an intentional crime was committed by two or more persons or units jointly.	1 = Yes; 0 = No
Intention	An INTENTION element refers to the subjective aspect of the defendant in an intentional crime as determined by the court.	Direct Intention/Indirect Intention
Negligence	A NEGLIGENCE element refers to the subjective aspect of the criminal in a negligent crime as determined by the court.	Carelessness/Overconfidence
Sufficient_evidence	A SUFFICIENT_EVIDENCE element refers to whether the court determined that evidence is sufficient.	1 = Yes; 0 = No
Beyond_reasonable_doubt	A BEYOND_REASONABLE_DOUBT element refers to whether the court determined that the evidence is strong enough to rule out any reasonable doubt.	1 = Yes; 0 = No
Criminal_pattern	A CRIMINAL_PATTERN element refers to the pattern of the crime.	Preparation/Discontinuation/Criminal Attempt/Consummation
Confession	A CONFESSION element refers to whether the court determined that the defendant confessed to authorities.	1 = Yes; 0 = No
Forgiveness	A FORGIVENESS element refers to whether the court determined that the defendant had obtained forgiveness from victims or their close relatives.	1 = Yes; 0 = No

Table 10: List of detailed element information (VI).

Element Name	Explanation	Value Type
Crime Characteristics (II)		
Settlement	A SETTLEMENT element refers to whether both parties reach a settlement.	1 = Yes; 0 = No
Subjective_malice	A SUBJECTIVE_MALICE element refers to the degree of the subjective malice of the defendant as determined by the court.	No or Low Subjective Malice/High Subjective Malice
Severity	A SEVERITY element refers to the level of severity of the crime as determined by the court.	Circumstances Clearly Minor/Circumstances Minor/Circumstances Serious or Execrable/Circumstances Very Serious or Execrable
Cause_death	A CAUSE_DEATH element refers to whether the court determined that whether any death occurred because of the crime.	1 = Yes; 0 = No
Death_number	A DEATH_NUMBER element refers to the total number of victims who died because of the crime according to the court. This element is annotated only when element VICTIM_DEATH is assigned a value of 1.	Numeric Value Calculated Based on the Content of Judicial Documents
Serious_injury	A SERIOUS_INJURY element refers to whether the court determined that serious injury of any victim is caused by the crime.	1 = Yes; 0 = No
Serious_injury_number	A SERIOUS_INJURY_NUMBER element refers to the total number of victims who were seriously injured by the crime according to the court. This element is annotated only when element SERIOUS_INJURY is assigned a value of 1.	Numeric Value Calculated Based on the Content of Judicial Documents
Heavier_punishment	A HEAVIER_PUNISHMENT element refers to the number of circumstances leading to heavier punishment within the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Lesser_punishment	A LESSER_PUNISHMENT element refers to the number of circumstances leading to lesser punishment within the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Aggravated_punishment	An AGGRAVATED_PUNISHMENT element refers to the number of circumstances leading to aggravated punishment above the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Mitigated_punishment	A MITIGATED_PUNISHMENT element refers to the number of circumstances leading to reduced punishment below the legally prescribed limits of punishment according to the court.	Numeric Value Calculated Based on the Content of Judicial Documents
Reoffending_danger	A REOFFENDING_DANGER element refers to whether the court determined that the defendant would likely to commit any crime again.	1 = Yes; 0 = No
Plead_guilty	A PLEAD_GUILTY element refers to whether the court determined that the defendant pled guilty.	1 = Yes; 0 = No
Punishment_acceptance	A PUNISHMENT_ACCEPTANCE element refers to whether the court determined that the defendant accepted punishment.	1 = Yes; 0 = No
Repentance	A REPENTANCE element refers to whether the court determined that the defendant had showed genuine repentance.	1 = Yes; 0 = No
Meritorious_service	A MERITORIOUS_SERVICE element refers to whether the court determined that the defendant performed meritorious service.	1 = Yes; 0 = No
Return	A RETURN element refers to whether the court determined that the defendant actively returned the property that the defendant acquired illegally or equivalent amount of money.	1 = Yes; 0 = No
Compensation	A COMPENSATION element refers to whether the court determined that the defendant actively made compensation to victims actively.	1 = Yes; 0 = No

Table 11: List of detailed element information (VII).

Element Name	Explanation	Value Type
Crime Characteristics (III)		
Pregnancy_or_miscarriage	A PREGNANCY_OR_MISCARRIAGE element refers to whether the court determined that the defendant was pregnant or had a miscarriage during prosecution or trial.	1 = Yes; 0 = No
Vicious_means	A VICIOUS_MEANS element refers to whether the court determined that the crime was conducted using vicious means.	1 = Yes; 0 = No
Personal_harm	A PERSONAL_HARM element refers to whether the court determined that the defendant posed a high risk of or caused personal harm.	1 = Yes; 0 = No
Legitimate_defense	A LEGITIMATE_DEFENSE element refers to whether whether the court determined that the act of the defendant was legitimate defense.	1 = Yes; 0 = No
Excessive_defense	An EXCESSIVE_DEFENSE element refers to whether the court determined that defendant's defense noticeably exceeded the necessary limits.	1 = Yes; 0 = No
Voluntary_Surrender	A VOLUNTARY_SURRENDER element refers to whether the court determined that the defendant voluntarily surrendered to the police and gave a true account of one's crime after committing it.	1 = Yes; 0 = No
Adverse_effect	An ADVERSE_EFFECT element refers to whether the court determined that the social effects caused by the defendant were significantly adverse.	1 = Yes; 0 = No
Minor_result	A MINOR_RESULT element refers to whether the court determined that the circumstances of the alleged conduct are obviously minor, causing no serious harm.	1 = Yes; 0 = No
Victim_fault	A VICTIM_FAULT element refers to whether the court determined that victim was in fault for the crime.	1 = Yes; 0 = No
Danger_prevention	An DANGER_PREVENTION element refers to whether the court determined that the defendant's act was a legitimate prevention of urgent danger.	1 = Yes; 0 = No
Excessive_danger_prevention	An EXCESSIVE_DANGER_PREVENTION element refers to whether the court determined that the defendant's act was a prevention of urgent danger that exceeded the necessary limits and caused undue harm.	1 = Yes; 0 = No
Principal_offender	A PRINCIPAL_OFFENDER element refers to whether the court determined that the defendant organized and leads a criminal group in conducting criminal activities or played a principal role in a joint crime. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Accomplice	An ACCOMPLICE element refers to whether the court determined that the defendant played a secondary or supplementary role in a joint crime. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Coerced_accomplice	A COERCED_ACCOMPLICE element refers to whether the court determined that the defendant was coerced to participate in a crime shall. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Abettor	An ABETTOR element refer to whether the court determined that the defendant was the one who instigated others to commit a crime and should be punished according to the role he played in the joint crime. This element is annotated only when JOINT_CRIME element is assigned a value of 1.	1 = Yes; 0 = No
Not_guilty	A NOT_GUILTY element refers to whether the court determined that the defendant was not guilty of a specific crime.	1 = Yes; 0 = No; Non-mention is treated as 0

Table 12: List of detailed element information (VIII).

Element Name	Explanation	Value Type
Crime Characteristics (IV)		
Exemption	An EXEMPTION element refers to whether the defendant is exempted from criminal punishment of a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Judgement_modified	A JUDGEMENT_MODIFIED element refers to whether the court determined that full revision and partial reversal with partial revision of sentence should be made of a specific crime.	1 = Yes; 0 = No; Non-mention is treated as 0
Judgement_upheld	A JUDGEMENT_UPHELD element refers to whether the court decided to uphold the previous judgment of a specific crime.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspension	A SUSPENSION element refers to whether the court determined a suspension of sentence for a specific crime.	1 = Yes; 0 = No
Suspension_term	A SUSPENSION_TERM element refers to the probation period for suspension for a specific crime as determined by the court. This element is annotated only when element SUSPENSION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Probation	A PROBATION element refers to whether the defendant was subject to probation for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Probation_term	A PROBATION_TERM element refers to the term of probation of a specific crime as determined by the court. This element is annotated only when element PROBATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Limited_incarceration	A LIMITED_INCARNATION element refers to whether the defendant was subject to limited incarceration for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Limited_incarceration_term	A LIMITED_INCARCERATION_TERM element refers to the term of limited incarceration as determined by the court for a specific crime of a defendant. This element is annotated only when element LIMITED_INCARNATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Fixed_imprison	A FIXED_IMPRISON element refers to whether the defendant was subject to fixed incarceration for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Fixed_imprison_term	A FIXED_IMPRISON_TERM element refers to the term of fixed-term imprisonment as determined by the court for a specific crime of a defendant. This element is annotated only when element FIXED_IMPRISON is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Life_imprison	A LIFE_IMPRISON element refers to whether the defendant was subject to life imprisonment for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Suspended_death_penalty	A SUSPENDED_DEATH_PENALTY element refers to whether the defendant was subject to a suspended death penalty for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Immediate_death_penalty	A IMMEDIATE_DEATH_PENALTY element refers to whether the defendant was subject to an immediate death penalty for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Political_deprivation	A POLITICAL_DEPRIVATION element refers to whether the defendant was subject to the deprivation of political rights for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Political_deprivation_term	A POLITICAL_DEPRIVATION_TERM element refers to the term of deprivation of political rights for a specific crime of a defendant as determined by the court. This element is annotated only when element POLITICAL_DEPRIVATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents

Table 13: List of detailed element information (IX).

Element Name	Explanation	Value Type
Crime Characteristics (V)		
Fine	A FINE element refers to whether the defendant was fined for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Fine_amount	A FINE_AMOUNT element refers to the amount of the fine of a specific crime as determined by the court. This element is annotated only when element FINE is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Partial_property_confiscation	A PARTIAL_PROPERTY_CONFISCATION element refers to whether the defendant was subject to confiscation of a part of his or her property for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Total_property_confiscation	A TOTAL_PROPERTY_CONFISCATION element refers to whether the defendant was subject to confiscation of all of his or her property for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscated_property_amount	A CONFISCATED_PROPERTY_AMOUNT element refers to the amount of confiscated property as determined by the court for a specific crime of a defendant. This element is annotated only when either element PARTIAL_PROPERTY_CONFISCATION or element TOTAL_PROPERTY_CONFISCATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Deportation	A DEPORTATION element refers to whether the defendant was subject to deportation for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Loss_compensation	A LOSS_COMPENSATION element refers to whether the defendant was ordered to make compensation for the economic loss for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Compensation_amount	A COMPENSATION_AMOUNT element refers to the amount of making compensation for the economic loss as determined by the court for a specific crime of a defendant. This element is annotated only when element LOSS_COMPENSATION is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Formal_apology	A FORMAL_APOLOGY element refers to whether the defendant was ordered to make a statement of repentance or formal apology for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Administrative_sanction	An ADMINISTRATIVE_SANCTION element refers to whether the defendant was subjected to administrative sanctions by the relevant department for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Occupational_prohibition	An OCCUPATIONAL_PROHIBITION element refers to whether the defendant was subjected to occupational prohibition for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscate_illegal_income	A CONFISCATE_ILLEGAL_INCOME element refers to whether the defendant was subjected to the confiscation of illegal income for a specific crime as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Confiscated_income_amount	A CONFISCATED_INCOME_AMOUNT element refers to the amount of confiscated income for a specific crime of the defendant as determined by the court. This element is annotated only when element CONFISCATE_ILLEGAL_INCOME is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_confiscation	An OTHER_CONFISCATION element refers to whether the defendant was subjected to the confiscation of other objects related to a specific crime of the defendant as determined by the court in the aggregated sentencing.	1 = Yes; 0 = No; Non-mention is treated as 0

Table 14: List of detailed element information (X).

Element Name	Explanation	Value Type
Crime Characteristics (VI)		
Return_illegal_gain	A RETURN_ILLEGAL_GAIN element refers to the order of returning illegal gains and compensations for a specific crime of the defendant as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0
Return_amount	A RETURN_AMOUNT element refers to the amount of returning illegal gains and compensations as ordered by the court for a specific crime of the defendant. This element is annotated only when element RETURN_ILLEGAL_GAIN is assigned a value of 1.	Extracted from the Specific Content of Judicial Documents
Other_nonpenalty_punishment	A OTHER_NONPENALTY_PUNISHMENT element refers to the order of returning illegal gains and compensations for a specific crime of the defendant as determined by the court.	1 = Yes; 0 = No; Non-mention is treated as 0

Table 15: List of detailed element information (XI).